图书情报工作
第63卷 第20期 2019年10月

数据驱动的微信用户信息行为时间特征研究*

■ 张大勇[1] 孔洪新[1] 许磊[2] 景东[1]
[1] 哈尔滨工业大学互动媒体设计与装备服务创新重点实验室 哈尔滨 150001
[2] 哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001

摘要：[**目的/意义**]相对于传统的信息行为分析，数据驱动的信息行为研究更注重数据的外在性与客观性，所得的结果能够更为全面地认识用户信息行为本质特征。[**方法/过程**]通过自行构建的APP实现对微信用户分享和阅读行为记录的采集，并对微信用户信息行为的时间特性进行系统的分析。[**结果/结论**]结果表明：微信用户日常信息行为存在显著的假日效应，但是在信息行为时间间隔分布上存在明显厚尾现象和很强的阵发性，预示着微信用户信息行为具有较高的复杂性和不确定性，无法对其产生过程实现有效的预测；此外，微信用户所分享的内容具有很强的时效性，多数内容在微信中能够得到及时的传播，但传播链长度显著受分享内容主题的影响。

关键词：微信 信息行为 复杂性 时间特征
分类号：G203
DOI：10.13266/j.issn.0252-3116.2019.20.012

1 引言

社交网络是集交互性、即时性和传播多样性为一体的在线服务平台，已成为众多用户满足信息、娱乐、人际关系等方面需求的网络聚集地，其复杂的信息行为深刻影响着国家信息安全、政治、经济、社会的稳定和人们的日常生活[1-2]。微信作为典型的即时通信类社交服务平台，其使用率已经连续多年位居社交类应用的榜首。截止到2018年9月微信全球月活用户达到10.8亿，用户日发送信息超过了450亿次，日发送音视频通话达到了4.1亿次。庞大的用户群体、多元化的交互方式和活跃的社交关系，使微信成为众多学者的研究重点[3-5]。如C. Gan[6]采用PLS-SEM模型对215份来自微信用户的在线调查数据进行了评估，研究表明享乐满足、社会支持和信息搜集均会对用户的点赞行为产生积极的影响；李嘉等[7]通过问卷调查发现功用满足、社交满足和享乐满足对微信用户的持续使用意愿有着显著的正向影响；李力和丁宁[8]则从移动社交类应用的生命周期出发，系统地梳理了以微信、微博为代表的移动社交平台用户信息行为的研究进展。

对人类信息行为研究是情报学关注的主要研究领域之一。早期的研究主要关注受众的阅读行为规律以及对文献的检索和使用[9]，其目的在于优化信息管理过程。在一般情况下，信息行为的研究范畴包括了用户在信息需求和动机基础上从事的需求认识与表达、信息查寻、信息选择、信息存储、信息吸收与利用以及信息加工、信息交互等活动[10]。随着社会的发展和新型信息技术的广泛应用，信息传播媒介的改变，对人类信息行为的研究内容也在不断地变化[11]。特别是搜索引擎和社交网络的广泛应用，为用户信息行为的定量分析提供了大量的鲜活数据，摆脱了原有研究中数据来源的限制。近年来，利用用户网络行为日志数据、以数据驱动为代表的用户信息行为研究成为了一个重点，它可以更为客观、系统地描述用户信息行为特征和行为规律[12-13]，并对用户的行为偏好[14]、需求认知[15]、行为习惯[16]等给出定量化的分析，从而为信息

* 本文系教育部人文社会科学基金面上项目"大数据驱动的社交媒体用户角色识别与链接预测研究"(项目编号：19YJA630106)和国家社会科学基金青年项目"社交媒体突发公共事件的协同应急机制研究"(项目编号：14CXW045)研究成果之一。
作者简介：张大勇(ORCID：0000-0001-9122-2220)，副教授，博士，E-mail：zhdy@hit.edu.cn；孔洪新(ORCID：0000-0001-8050-1080)，助理工程师；许磊(ORCID：0000-0002-1112-8398)，高级工程师，博士；景东(ORCID：0000-0001-9550-9595)，工程师，博士研究生。
收稿日期：2019-01-24 修回日期：2019-04-16 本文起止页码：103-111 本文责任编辑：徐健

张大勇,孔洪新,许磊,等.数据驱动的微信用户信息行为时间特征研究[J].图书情报工作,2019,63(20):103-111.

系统构建精准化、个性化的信息检索与推荐服务提供帮助[17]。

以数据驱动的人类信息行为统计特征研究始于2005年美国东北大学A. L. Barabási教授[18]对学校3 188位用户收发电子邮件的数据分析,研究表明用户行为在时间分布上存在对泊松过程的偏离,而这是传统研究范式所无法解释的;此后,在对网络电影点播[19]、网页浏览[20]、网络音乐下载[21]、社区互动[22]等行为研究中均发现有类似的非泊松特征,表明用户信息行为在时空分布上非均匀、不稳定可能是一种普遍存在的现象。为此,K. I. Goh和A. L. Baraasi对用户在时空方面表现出的异常行为提供了检测方法[23]。这种通过人类动力学揭示用户日常的信息应用模式,不同于传统的定性或半定量化分析,而是通过无干扰的客观数据,从一个外在观察者视角给出分析结果,其遵循的是"观察-数据获取与分析-统计规律挖掘-建模再现数据规律"的过程[24]。

综上所述,信息行为的研究范畴十分广泛,涉及到从信息需求认识与表达、查寻、选择、存储、加工到用户间交互等全过程;对用户信息行为的研究方法从定性或半定量的描述逐渐转移到以数据驱动的信息行为挖掘。数据驱动的信息行为研究不同于传统的研究,其更多地是关注于用户行为与各类信息系统平台及信息本身自然交互过程中留下来的行为轨迹,数据是外在的、非介入性的、且客观的,对于更全面地认识和优化用户信息行为,了解用户信息行为特征和习惯,为揭示人类信息行为的复杂性提供了新的研究思路和实施途径。然而由于微信不提供开放的数据接口,无法获取微信个体用户的行为信息,现有的研究多以访谈和发放问卷形式了解微信用户的行为特征与规律以及影响因素[25-28]。但是采用访谈和问卷调查具有很大的局限性,一是难以对用户行为时空分布特征给予客观准确的描述;二是无法对用户行为关联度以及用户间信息分享轨迹给予量化分析,这些都极大地限制了对微信用户信息行为的研究。为此,笔者通过自行构建的移动APP平台采集、分析微信用户信息行为记录,既可以发现用户个体行为统计特征,也可将孤立的个体相互关联起来,分析用户间信息分享轨迹,从而实现宏观趋势的预测。

2 数据获取与处理

沟通交流、关注新闻热点、获取感兴趣的内容和分享信息是人们使用社交网络的主要目的。而快速增长的用户数量和海量的用户行为数据也给社交网络带来了无限的商机。如Facebook在共享的技术架构基础上对数据传输、存储和数据处理系统进行优化,新浪微博也采用相似的技术架构并开放了API接口满足了移动终端用户多元化的产品需求。虽然微信也为公众平台开发提供技术支持,但是作为典型熟人社交工具,微信更加注重用户的私密性和信息传递的准确性,从产品设计上就限制了数据的开放程度,使之难以实现对微信普通用户行为信息的大数据采集。为此,笔者开发了APP"在工大"(@HIT),并在APP文章中植入阅读日志跟踪程序,使用者将APP内文章分享到微信朋友圈、微信好友、微信群后,该文章的阅读和再次分享的日志将被记录在服务器中,跟踪程序将记录分享用户ID、阅读用户ID、分享时间、阅读时间、分享场景、文章ID、文章主题、主题类别,由此实现对每一个用户分享、阅读日志的采集。

从2015年8月31日到2018年1月5日,笔者通过上述采集方式共获取了247 711条微信用户分享数据,涉及200 023位微信用户,采集文章总数为2 128篇。为了保障用户的个人隐私,所采集的数据均进行了脱敏处理,而且采集的数据仅作为科学研究使用,不作为任何商业用途。分享者ID和阅读者ID分别为数据脱敏后重新编号;数据来源表示分享文章的来源,包括移动终端(APP)、网页(Device)、微信(WX);阅读类型(to_type)专指文章被分享到微信中,并被微信用户阅读;分享时间(share time)和阅读时间(time stamp)分别用时间戳形式标注;分享场景是指用户分享的微信去向,包括好友消息(singlemessage)、朋友圈(timeline)、微信群聊(groupmessage);文章ID是分享文章的唯一标记,文章主题(title)是指分享者提供的文章标题,主题类别(subscribe_name)是指被分享文章的出处。具体见图1。

3 微信用户信息行为时间特性分析

人类行为时间特性关注的是人多次从事某一特定事件在时间轴上所表现出来的统计规律。笔者将从两个方面分析微信用户信息行为的时间特性:一是微信用户日常信息行为时间分布;二是微信用户连续信息行为间隔时间分布。

3.1 用户日常信息行为时间分布

微信用户日常信息行为包括分享行为(share)、阅读行为(stamp)和转发行为(reshare)。笔者将对微信用户上述3种行为的时间特征进行分析,其中:分享时

1	id	from_id	to_id	from_type	to_type	share_time	time_stamp	article_id	link_type	title	subscribe_name
2	66329	oRAnzt5X-	oRAnzt6pj	wx	wx	1445693616	1445695186	12839	singlemess	黑龙江省伊春市第	招聘
3	66330	oRAnzt7NB	oRAnzt_V0	wx	wx	1445694658	1445695186	12839	timeline	黑龙江省伊春市第	招聘
4	66331	oRAnzt5wq	oRAnzt6eF	wx	wx	1445688187	1445695187	12839	timeline	黑龙江省伊春市第	招聘
5	66332	oRAnzt9ze	oRAnzt2z-	wx	wx	1445694276	1445695190	12839	timeline	黑龙江省伊春市第	招聘
6	66333	oRAnzt1-v	oRAnzt6TX	wx	wx	1445694946	1445695190	12839	timeline	黑龙江省伊春市第	招聘
7	66334	oRAnzt3MP	oRAnzt7_v	wx	wx	1445694409	1445695191	12839	timeline	黑龙江省伊春市第	招聘
8	66335	oRAnzt9Kc	oRAnzt1Gc	wx	wx	1445694854	1445695201	12839	timeline	黑龙江省伊春市第	招聘
9	66336	oRAnzt-0Z	oRAnzt8Tb	wx	wx	1445683213	1445695210	12839	timeline	黑龙江省伊春市第	招聘
10	66337	oRAnzt2IZ	oRAnzt8ea	wx	wx	1445693804	1445695211	12839	timeline	黑龙江省伊春市第	招聘
11	66338	oRAnzt2f8	oRAnzt9oP	wx	wx	1445695152	1445695211	12839	groupmessa	黑龙江省伊春市第	招聘
12	66339	oRAnzt300	oRAnzt14z	wx	wx	1445693529	1445695215	12839	groupmessa	黑龙江省伊春市第	招聘
13	66340	oRAnzt8Fv_	oRAnzt6hn	wx	wx	1445694295	1445695221	12839	timeline	黑龙江省伊春市第	招聘
14	66341	oRAnzt4ED	oRAnzt2r1	wx	wx	1445694424	1445695228	12839	timeline	黑龙江省伊春市第	招聘
15	66342	oRAnzt_aF	oRAnztx2Y	wx	wx	1445694286	1445695231	12839	timeline	黑龙江省伊春市第	招聘
16	66343	oRAnzt0nH	oRAnzt4Rm	wx	wx	1445694310	1445695235	12839	timeline	黑龙江省伊春市第	招聘
17	66344	oRAnzt3Iy	oRAnzt9D8	wx	wx	1445688555	1445695260	12839	timeline	黑龙江省伊春市第	招聘
18	66345	oRAnzt47u	oRAnzt8ID	wx	wx	1445689063	1445695263	12839	timeline	黑龙江省伊春市第	招聘
19	66346	oRAnzt6Xi	oRAnzt4Kb	wx	wx	1445681900	1445695267	12839	timeline	黑龙江省伊春市第	招聘
20	66347	oRAnzt0nH	oRAnzt9nS	wx	wx	1445694310	1445695278	12839	timeline	黑龙江省伊春市第	招聘
21	66348	oRAnztxv-	oRAnzt0n4	wx	wx	1445694816	1445695280	12839	timeline	黑龙江省伊春市第	招聘
22	66349	oRAnzt3MP	oRAnztxdn	wx	wx	1445694409	1445695280	12839	timeline	黑龙江省伊春市第	招聘
23	66350	oRAnzt8Lc	oRAnzt9dh	wx	wx	1445692661	1445695291	12839	timeline	黑龙江省伊春市第	招聘
24	66351	oRAnzt-Vc	oRAnzt9GH	wx	wx	1445694702	1445695292	12839	groupmessa	黑龙江省伊春市第	招聘
25	66352	oRAnzt0nH	oRAnzt9It	wx	wx	1445694310	1445695296	12839	timeline	黑龙江省伊春市第	招聘
26	66353	oRAnzt9uL	oRAnzt5rr-	wx	wx	1445676376	1445695303	12839	timeline	黑龙江省伊春市第	招聘
27	66354	oRAnzt7NB	oRAnztxp_	wx	wx	1445694658	1445695304	12839	timeline	黑龙江省伊春市第	招聘
28	66355	oRAnzt-1S	oRAnzt54R	wx	wx	1445692805	1445695307	12839	timeline	黑龙江省伊春市第	招聘
29	66356	oRAnztxv-	oRAnzt2dl	wx	wx	1445694816	1445695308	12839	timeline	黑龙江省伊春市第	招聘
30	66357	oRAnzt4ED	oRAnzt8w-	wx	wx	1445694424	1445695312	12839	timeline	黑龙江省伊春市第	招聘

图 1　所采集样本的数据格式

间是指用户通过移动或固定终端从 APP、网页或微信中获取相关文章并将其转发到微信好友、朋友圈或微信群中的时间；阅读时间是指用户通过微信好友、朋友圈或微信群中阅读文章的时间；转发时间是指用户通过微信好友、朋友圈或微信群中阅读文章后再次转发的时间。

首先对微信用户日常使用习惯进行分析。由于数据时间跨度较大，为确定每天微信分享行为、阅读行为和转发行为的时间规律，笔者对原时间数据进行了抽取，分别按照小时和星期读取记录。图 2 标注了 24 小时内微信用户分享、阅读和转发行为的时间分布。从微信用户日常使用时间上看，用户的活跃度共经历两次高峰和一次低谷：从早晨 4 点起，微信用户使用量开始上升，到上午 9 点 – 10 点达到高峰，随后在 10 点 – 11 点出现小幅下滑，在 12 点午休时间有所回升，其后 3 个小时内有所下滑，下午 15 点之后微信用户使用量平缓上升，在晚上 22 点左右出现全天另一次使用高峰，此后开始快速下滑，在凌晨 3 点左右出现最低点。从微信用户日常使用时间分布来看，微信用户的使用高峰不再局限于休息时间，在用户处于工作或学习状态下，微信的使用量不降反升，说明微信已经深深地介入到用户日常工作和学习中。再者，微信用户活跃度会在节假日明显提升，表明微信的使用存在显著的周末效应。如图 3 所示，周六和周日微信用户使用量明显高于日常工作时间用户的使用量，而在整个样本采集时间段内，节假日微信用户分享行为占分享总量的 46.7%，阅读行为占阅读总量的 60.6%，转发行为占总量的 56.3%。这说明，微信作为一类社交应用，不仅满足着日常用户信息需求，而且正在挤占用户整体休息时间，表明用户在对外交往中对微信的依赖度得到显著提升。但是另一方面也反映出，微信过多地占用了用户的休息时间，对网络的依赖势必将减少用户现实人际互动的时间和机会，导致用户人际交往能力的下降和情感的缺失。

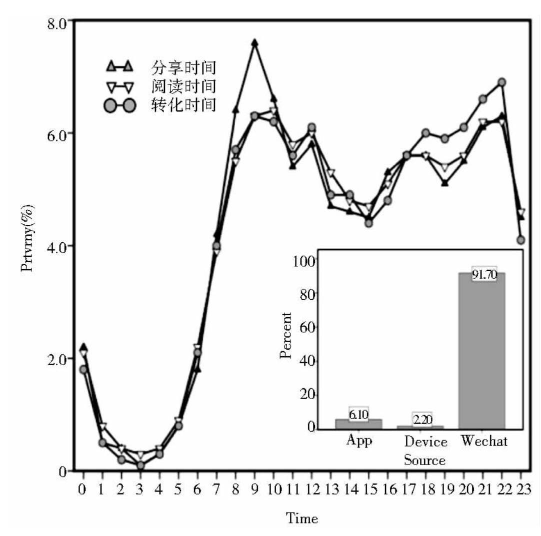


图 2　微信用户日常使用时间分布

3.2　用户连续信息行为时间间隔分布

长久以来，对人类行为时间特征的研究多停留在

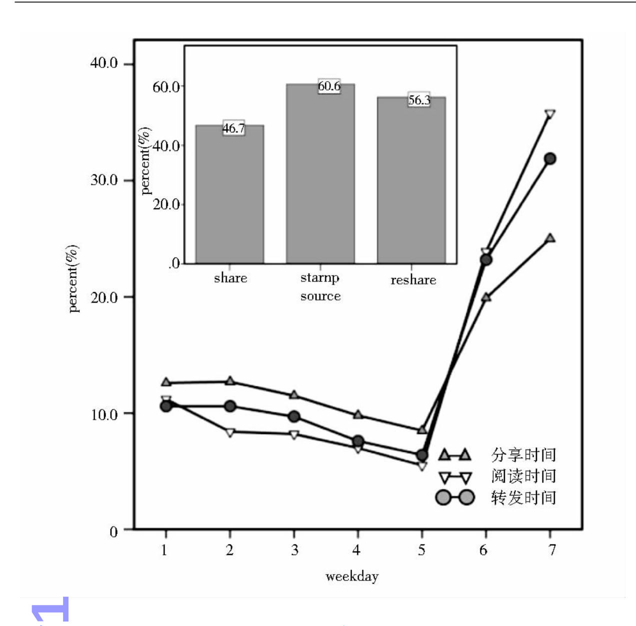


图 3　微信用户周末效应

用泊松分布来描述,而泊松过程可以视为一个具有负指数间隔的计数过程,因此,两次连续活动的间隔时间是近似均匀的,长时间内或短时间内同类行为密集出现的概率几乎为 0。然而,最近几年的研究显示,包括通讯、购物、交易等在内的人类许多行为在时间分布呈现了对泊松过程的偏离[19-22],而最为典型的非泊松过程是行为时间序列存在阵发性和记忆性。笔者根据样本中每一位用户行为发生时间构建用户行为时间序列,分析用户连续信息行为的时间间隔分布特征,以及从信息源首次分享后所经历的阅读、转发过程的时间间隔,判断信息在微信中的时效性。

首先,计算微信用户连续信息行为的时间间隔分布,并依此判断统计区间内微信用户信息行为是否存在阵发性和记忆性。其中,阵发性是指将用户行为作为一个事件,其发生频率是不均匀的,有时候很多事件会在较短的时间内集中发生,而有时候事件发生的间隔期又会很长;如果在一个行为发生的时间序列中,长的时间间隔后面也是长的时间间隔,或短的时间间隔后面跟着一个短的时间间隔,则认为该时间间隔序列存在记忆性。对于人类行为活动中表现出来的阵发性和记忆性,一般可采用 K. I. Goh 和 A. Barabási 提出的异常行为检测指标进行分析[23]。阵发性检测系数为 $B=\frac{\sigma_\tau - m_\tau}{\sigma_\tau + m_\tau}$,关注于时间间隔序列标准差 σ_τ 和均值 m_τ 的统计,$B\in(-1,1)$。如果 $B>0$ 表明该时间序列具有阵发效应,而且 B 越接近 1,则时间间隔分布越不均匀,阵发效应越强;$B=0$ 表示时间序列满足泊松分布,均值与标准差相等,是一个无倾向的序列。

如图 4(a)所示,从总体上看,微信用户信息行为发间隔时间分布存在明显偏离指数分布的厚尾现象,其分布曲线 $P(t)=1.004x^{-1.532}$,$R^2=-.964$,曲线符合幂律分布,显示行为产生时间间隔整体呈现极度不均匀,多数行为发生的时间间隔较为短(2 分钟之内),约占行为发生总数的 83.78%,但是也有少部分信息行为发生的时间间隔较长,时间间隔超过 1 个小时约占行为发生总数的 0.58%,而最长时间间隔高达 5 826 分钟。从用户个体角度,用户个体信息行为发生次数也是极度不均衡的,如图 4(b)所示,样本期间内用户个体信息行为发生次数超过 50 次的仅有 10 位用户,84.31% 用户统计期内信息行为次数仅为 1 次,最多行为发生数为 573 次。无论是从信息行为的间隔时间分布,还是从用户个体信息行为的发生次数来看,均呈现与泊松分布的钟型曲线完全不同的厚尾形态,表明微信用户在小范围内(有限的参与个体和有限的间隔时间)存在高频信息活动,而这些高频信息活动对于促进信息传播和提高传播效率起到了关键作用,是微信日常管理中需要着重关注的。

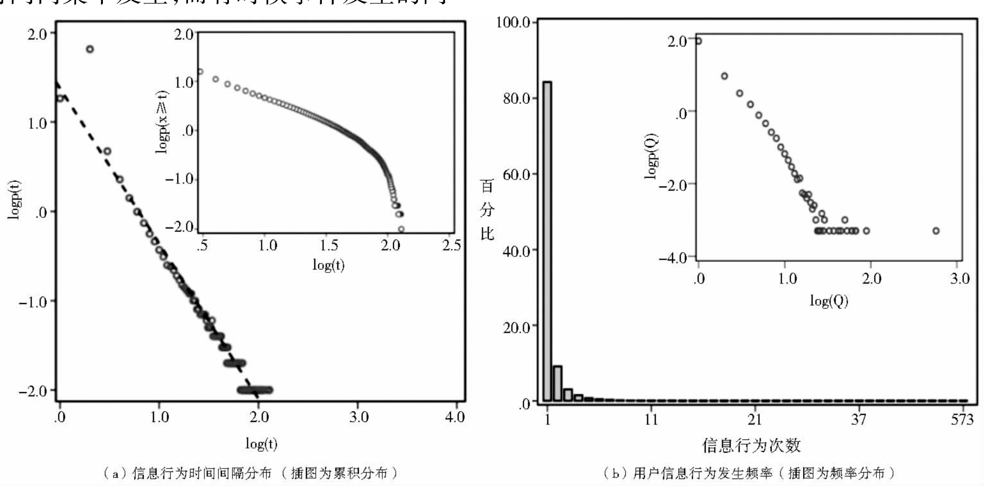


图 4　微信用户连续信息行为时间间隔

图 5 列示了从样本采集期第 1 分钟到样本采集第 100 小时内不同时间粒度下用户信息行为的发生密度。以小时作为间隔时间标注为例，从样本采集期开始 1 小时内发生了一次信息行为，此后经过了近 24 小时的静默期，在间隔时间为 25 小时和 26 小时内连续发生了两次信息行为，然后进入 7 个小时的静默期，在间隔时间为 34 小时、35 小时和 36 小时内连续发生了 3 次信息行为，此后在经历一个漫长的空档期后进入行为发生的密集期。从总体上看，大部分行为发生间隔时间和其平均值之间差异较大，以分钟为单位，微信用户信息行为发生的间隔时间分布的标准差 σ_τ 为 145.39，平均值 m_τ 仅为 5.32，阵发系数 $B=0.929$，接近 1，表明微信用户信息行为序列极不均衡，存在很强的阵发性。从分享场景上看（见表 1），在较小的时间粒度下，朋友圈（timeline）和微信群聊（groupmessage）中用户信息行为的阵发系数要高于好友消息（singlemessage），而随着时间粒度的增大，阵发系数有减弱趋势，特别是在朋友圈（timeline）中表现的更为显著。

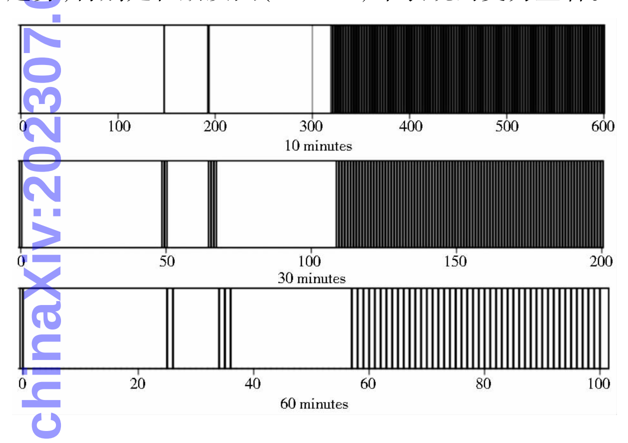


图 5　不同时间粒度下信息行为发生间隔

表 1　不同场景和时间粒度下信息行为阵发性检测

类型	时间粒度	平均值	标准差	阵发系数
Timeline	1 minute	6.76	238.39	0.94
	10 minute	1.57	23.83	0.88
	30 minute	1.19	7.94	0.74
	60 minute	1.09	3.96	0.57
Groupmessage	1 minute	35.59	928.16	0.93
	10 minute	4.43	92.80	0.91
	30 minute	2.13	30.93	0.87
	60 minute	1.55	15.45	0.82
Singlemessage	1 minute	84.57	1 485.15	0.89
	10 minute	9.20	148.50	0.88
	30 minute	3.66	49.49	0.86
	60 minute	2.28	24.73	0.83

记忆性则关注于连续两组行为时间序列之间的关联关系，K. I. Goh 和 A. Barabási 在文献[23]中提出了记忆相关系数 $M=\frac{1}{n_\tau-1}\sum_{i=1}^{n_\tau}\frac{(\tau_i-m_1)(\tau_{i+1}-m_2)}{\sigma_1\sigma_2}$，其中 n_τ 是第 τ 个用户行为时间间隔序列的数目，m_1 和 m_2 是计算过程所抽取的两个样本 τ_i 和 t_{i+1} 的平均值，σ_1 和 σ_2 是与之对应的标准差，如果 M 值为正，则时间序列具有记忆效应，M 值为负则为反记忆效应。按照记忆相关系数的定义，需要对每位用户分享行为时间序列进行分组，而每组 n_τ 个元素则意味着需要有 $n_\tau+1$ 个分享行为产生。如果选择每组抽取样本序列数目过少，则很难保证用户行为的连贯性。为此笔者选取信息行为频率较高的前 50 位用户作为样本，样本用户信息行为发生数最高为 573 次，最低为 19 次。经计算，样本用户记忆相关系数 M 最高值为 0.14，最低值为 -0.13，平均 $M=-0.03$，近似等于 0，表明在统计期内多数样本个体行为不存在记忆效应。如图 6 所示，图中每个点代表满足条件样本个体，除 1 个用户信息行为次数超过 500 次之外，样本用户信息行为次数都在 100 次以下；样本用户记忆系数基本上围绕 0 值成对称分布，而样本用户阵发性系数均大于 0，均值等于 0.4。表明在样本统计期内，高频微信用户的信息行为在时间间隔分布上存在显著的阵发性，但不存在记忆效应。

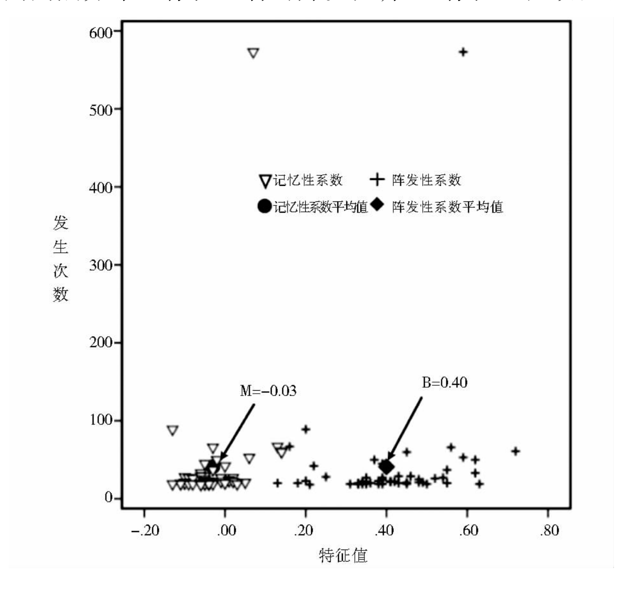


图 6　高频用户行为时间间隔特性检测

微信中的信息传播是依托人际关系链带动的信息传播链，信息只有通过用户层层分享和转发才能产生传播效果。分析微信用户之间的分享与阅读、阅读与转发行为之间时间间隔，将有助于了解微信中信息传播的时效性。通常情况下，信息从信息源发送后信息经历的分享、阅读的时间间隔越短，则信息被采纳的越

及时，信息的时效性越强；而信息传播链条越长，则反映信息被转发的层级越多，信息有效到达率和用户的参与深度都得到了强化。笔者将信息的时效性定义为 $C_k = \frac{1}{w_k}\sum_{i \neq j} a_{ij}^k$，其中 a_{ij}^k 表示针对第 k 篇文章用户 i 与用户 j 之间阅读与分享的时间间隔，w_k 表示第 k 篇文章被阅读的次数，则平均时效性为 $C' = \frac{1}{n}\sum_{k=1}^{n}(w_k C_k)$；信息的传播链长度可以定义为 $D_k = \max_{i \neq j} d_{ij}^k$，其中 d_{ij}^k 表示针对第 k 篇文章所构建的传播链中任意两个用户 i 与 j 之间测地线长度，则平均传播链长度 $D' = \frac{1}{n}\sum_{k=1}^{n} D_k$；$k \in [1, n]$，$n = 2128$。

微信是典型的即时类通信工具，实现两人或多人之间的高效沟通是人们使用微信的主要目的，分享与阅读之间的时间间隔能够准确地反映信息在微信中的传播时效。如图 7(a)所示，各篇文章分享与阅读之间的时间间隔分布是极不均衡的：从每篇文章分享与阅读的平均间隔时间上看，在 1 分钟之内被阅读的比重为 13.44%；而从分享到被阅读间隔时间超过 24 小时的约占文章总数的 6.06%；文章的平均时效性集中在 1－120 分钟区间内，约占文章总数的 62.5%。分享与阅读的平均间隔时间反映的各篇文章总体的时效性，但是受微信用户使用习惯和主题偏好的影响，每位用户对文章利用程度是存在很大差异的，由此导致每篇文章分享与阅读的间隔时间分布极不均衡，为了准确地反映信息传播效率，笔者选择测量每篇文章分享与阅读间最短时间差，如图 7(b)所示，最短时间差在一分钟之内的文章比重为 37.8%，30 分钟之内被阅读的文章比重高达 81.95%，表明绝大多数信息在微信中能够得到及时的传播，而且微信用户对信息的传递和利用的效率也是比较高的。这一点从各篇文章的传播链长度得以进一步验证，如图 8 所示，传播链长度累积分布满足指数为 2.22 的幂律分布，这意味着绝大多数文章被持续转发的次数很低，仅存在少数文章能够得到众多微信用户持续关注和转发。

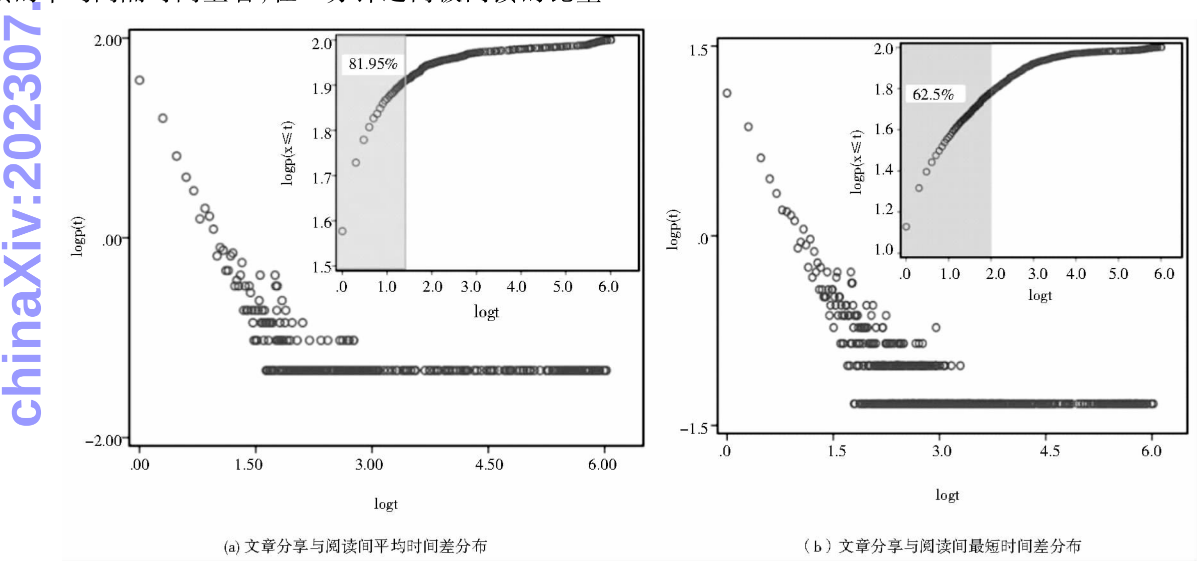


图 7　文章时效性分析

事实上也是如此，在总数为 2 128 篇的文章中，单篇传播链最长为 51，传播链最短为 1，平均分享路径长度为 10。传播链长度超过 10 的仅占文章总数的 0.47%，但却占到信息行为总次数的 80.62%；而传播链长度为 1 的文章占到文章总数的 88.44%，但却仅占信息行为总次数的 0.36%。传播链长度反映了用户对分享内容的持续关注程度，传播链越长则表明该内容在微信中被转发的次数越高，微信用户的参与度越高。上述结果意味着绝大多数文章并没有得到微信用户的普遍关注，也没有产生转发价值。由于“在工大”(@HIT)使用者多为在校师生，在微信中分享的文章主题涉及校园公告、学校新闻、校园招聘和兼职、校园美食、旅游攻略等，在传播链长度超过 10 的 14 篇文章中，涉及校园公告有 6 篇，学校新闻有 4 篇，校园招聘有 4 篇，都是与学生在校学习、生活紧密相关的主题。

4　结论

微信作为最具影响力社交应用，对其使用行为的研究一直都是研究的热点。笔者以微信作为研究对象，在技术路线上，遵从人类动力学“观察－数据获取

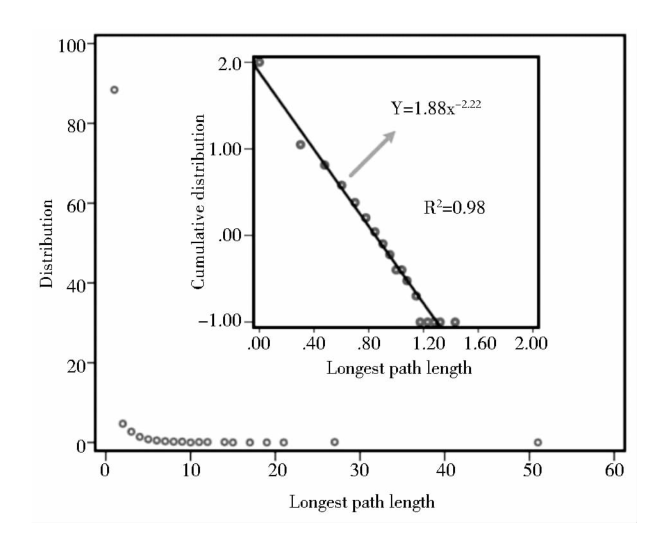


图 8 微信中单篇文章传播链长度

与分析－规律挖掘与再现"过程。首先通过自行构建的移动 APP 平台实现了对微信用户分享和阅读行为记录的采集,采集的数据包括时间信息、分享关系、分享场景、分享主题。而对微信用户信息行为的时间特性的揭示主要从两个方面展开:①对微信用户日常信息行为时间分布特征的分析,研究表明微信用户日常使用高峰不再局限于休息时间,而已经深深地介入到用户日常生活和工作中,并且微信用户活跃度在节假日会明显提升,存在显著的假日效应。②对微信用户信息行为间隔时间分布特征的分析,研究表明微信用户连续信息行为之间的时间间隔分布存在明显厚尾现象和很强的阵发性,并且时间粒度越小,阵发性越强,在微信朋友圈中表现的尤为显著;而通过对高频用户的连续信息行为间的关联性分析表明,统计期内样本用户信息行为不存在显著的记忆效应;微信用户表现出的强阵发性和无记忆性特征,意味着微信中用户信息行为具有较高的复杂性和不确定性,对其产生过程很难实现有效的预测。此外,通过文章时效性分析表明,绝大多数信息在微信中能够得到及时的传播,并且微信用户对信息的传递和利用的效率较高,但是受分享内容主题的影响,多数内容没有得到用户的持续关注,也没有产生转发价值。

参考文献:

[1] OH O, EOM C, RAO H R. Role of social media in social change: an analysis of collective sense making during the 2011 Egypt revolution[J]. Information systems research, 2015, 26(1): 210－223.
[2] 李栋,徐志明,李生,等. 在线社会网络中信息扩散[J]. 计算机学报,2014,37 (1): 189－206.
[3] LIU L, QU B, CHEN B, et al. Modelling of information diffusion on social networks with applications to WeChat[J]. Physica A, 2018, 496: 318－329.
[4] ZHANG Y X, LI Z Q, GAO C L, et al. Mobile social big data: WeChat moments dataset, network applications, and opportunities[J]. IEEE network, 2018, 32(3): 146－153.
[5] XU X, YAO Z, SUN Q. Social media environments effect on perceived interactivity: an empirical investigation from WeChat moments[J]. Online information review, 2019, 43(2): 239－255.
[6] GAN C. Understanding WeChat users' liking behavior: an empirical study in China[J]. Computers in human behavior, 2017, 68: 30－39.
[7] 李嘉,任嘉莉,刘璇,等. 微信公众平台的用户持续使用意愿研究[J]. 情报科学, 2016 (10): 26－33,55.
[8] 李力,丁宁. 国内外移动社交类应用用户信息行为研究进展[J]. 图书情报工作, 2015, 59 (10): 137－144.
[9] BELKIN N. The cognitive viewpoint in information science[J]. Journal of information science, 1990, 16(1): 11－15.
[10] 王艳,邓小昭. 网络用户信息行为基本问题探讨[J]. 图书情报工作, 2009, 53(16): 35－39.
[11] 叶平浩,张李义. 2002－2011 年国内外信息行为研究对比分析[J]. 情报科学,2013(4): 149－154.
[12] KOSINSKI M, BACHRACHY, KOHLI P, et al. Manifestations of user personality in website choice and behaviour on online social networks[J]. Machine learn, 2014, 95(3): 357－380.
[13] PHAM T, SHERIDAN P, SHIMODAIRA H. Joint estimation of preferential attachment and node fitness in growing complex networks[J]. Scientific reports, 2016, 6: 1－13.
[14] ZHANG Z, LIU L, WANG H, et al. Collective behavior learning by differentiating personal preference from peer influence[J]. Knowledge-based systems, 2018, 159: 233－243.
[15] GARG S, KAUR K, KUMAR N, et al. Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: A social multimedia perspective [J]. IEEE transactions on multimedia, 2019, 21(3): 566－578.
[16] SEOL S, LEE H, YU J, et al. Continuance usage of corporate SNS pages: a communicative ecology perspective[J]. Information & management, 2016, 53(6): 740－751.
[17] 李月琳,章小童. 数据驱动的信息行为研究的回顾与展望[J]. 信息资源管理学报, 2018(2): 13－27.
[18] BARABÁSI A-L. The origin of bursts and heavy tails in human dynamics[J]. Nature, 2005, 435 (7039): 207－211.
[19] DEZSÖ Z, ALMAAS E, LUKÁCS A, et al. Dynamics of information access on the web[J]. Physical review E, 2006, 73(6): 066132.
[20] ZHOU T, KIET H A T, KIM B J, et al. Role of activity in human dynamics[J]. Europhysics letters, 2008, 82(2): 28002.
[21] HU H B, HAN D Y. Empirical analysis of individual popularity and

activity on an online music service system[J]. Physica A, 2008, 387(23):5916－5921.

[22] GRABOWSKI A, KRUSZEWSKA N, KOSIŃSKI R A. Dynamic phenomena and human activity in an artificial society[J]. Physical Review E,2008,78(6): 066110.

[23] GOH K I, BARABÁSI A-L. Burstiness and memory in complex systems[J]. Europhysics letters,2008, 81(4):48002.

[24] 周涛,韩筱璞,闫小勇,等. 人类行为时空特性的统计力学[J]. 电子科技大学学报, 2013, 42(4):481－539.

[25] 薛杨,许正良. 微信营销环境下用户信息行为影响因素分析与模型构建－基于沉浸理论的视角[J]. 情报理论与实践,2016,(6):104－109.

[26] CHEN L, CHIN F G, SUN Y, et al. Integrating guanxi into technology acceptance: an empirical investigation of WeChat[J]. Telematics and informatics,2017,34(7):1125－1142.

[27] 吕亚兰,黄成,周虎. 微信平台用户健康信息行为及其影响因素研究[J]. 医学信息学杂志,2018,39(3):77－80.

[28] WANG G, ZHANG W, ZENG R X. WeChat use intensity and social support: the moderating effect of motivators for WeChat use [J]. Computers in human behavior,2019,91:244－251.

作者贡献说明：

张大勇:提出研究思路,设计研究方案并撰写论文;

孔洪新:参与数据采集和处理;

许磊:参与部分论文撰写;

景东:参与论文的修订。

Temporal Characteristics of Wechat Users' Information Behavior Based on Data-driven Approach

Zhang Dayong[1] Kong Hongxin[1] Xu Lei[2] Jing Dong[1]

[1] Key Laboratory of Interactive Media Design and Equipment Services Innovation, Harbin Institute of Technology, Harbin 150001

[2] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001

Abstract: [**Purpose/significance**] Compared with traditional information behavior approaches, the research on information behavior based on the data-driven approach pays more attention to the externality and objectivity of data, and the testing results can be more comprehensive understanding of user information behavior characteristics. [**Method/process**] This paper realizes the collection of Wechat users' sharing and reading behavior records through a self-built APP, and systematically analyses the temporal characteristics of Wechat users' information behavior. [**Result/conclusion**] The results show that the daily information behavior of Wechat users has significant holiday effect, but there are a obvious fat-tail phenomenon and strong burstiness effect in the time interval distribution of information behavior, which indicate that the information behavior of Wechat users has high complexity and uncertainty, and can not effectively predict its generating process; on the other hand, when the contents shared by Wechat users have very strong time-effectiveness, the most contents can be timely disseminated in Wechat, but the length of the dissemination chain is significantly affected by the theme of the shared contents. This study provides a reference for revealing the complexity of human information behavior.

Keywords: Wechat information behavior complexity temporal characteristics